2. Strumenti utili per la realizzazione di una IM

In questo capitolo analizzeremo i vari componenti necessari per sviluppare un'interfaccia multimodale.

2.1 Speech Engine

Dopo numerosi studi , il mercato multimediale si è attivato molto nel proporre nuove tecnologie per la progettazione delle interfacce; tra tutte le proposte, le più interessanti sono sicuramente quelle che impiegano il riconoscimento e la sintesi vocale per consentire l'interazione tra gli utenti ed il computer. La possibilità di convertire la voce in comandi, dati e testo può ridurre l'interfaccia uomo/macchina a pochi elementi visuali.

Anche se tali tecnologie sono presenti sul mercato da diversi anni, il loro uso all'interno delle applicazioni software è rimasto limitato a qualche grande prodotto a causa dei numerosi problemi di "comunicazione" che si incontravano durante il riconoscimento della voce.

Uno speech engine (spesso chiamato Text-to-Speech in virtù della sua capacità di tramutare il testo scritto in audio) si occupa della connettività audio e dell'utilizzo dei motori di elaborazione vocale ; esso può essere completamente software oppure supportato da hardware dedicato (esempio schede DSP).

Uno speech engine è composto da due parti: una detta front-end , l'altra back-end.

La parte front-end prende l'ingresso in forma di testo e fornisce in uscita una rappresentazione linguistica simbolica mentre la parte back-end prende la rappresentazione linguistica simbolica in ingresso e fornisce in uscita la voce sintetizzata.

Esistono diverse tecniche per realizzare le operazioni sopra descritte, ma in linea generale tutte seguono i seguenti passi:

- il front-end prende il testo grezzo e converte simboli come numeri e abbreviazioni in termini specifici equivalenti. Questo processo è spesso chiamato text-normalization o pre-processing. Quindi assegna le trascrizioni fonetiche a ciascuna parola e divide il testo in periodi e frasi;
- il back-end prende in ingresso l'uscita del front-end e produce in uscita la rappresentazione sonora del testo inserito;

Un engine non deve essere necessariamente locale, cioè sulla macchina su cui si sta operando; infatti è possibile abilitarlo remotamente su di un server ed interagire con esso attraverso una rete.
I passi da compiere per utilizzare uno speech engine all'interno di una applicazione sono:

1. identificare le funzionalità necessarie dall'applicazione, come ad esempio la lingua;
2. localizzare e creare un engine che possieda le caratteristiche richieste;
3. rendere disponibili le risorse per l'engine e configurarlo.

Naturalmente dopo l'uso è necessario rilasciare le risorse impiegate per consentirne l'uso in altre applicazioni.

Un'applicazione deve essere sviluppata in modo che possa identificare automaticamente le proprietà da richiedere a uno speech engine, ad esempio essere in grado di riconoscere il parlato nella lingua locale e di riprodurre un testo usando una voce femminile. Le proprietà di un engine possono variare in funzione della sua progettazione e di come è stato implementato: è possibile fare in modo che sia in grado di operare in diverse modalità, ciascuna delle quali identificata da un unico set di proprietà incapsulate in speciali oggetti detti "mode descriptor". In questo modo, l'applicazione è in grado di selezionare un dato engine basandosi proprio sulle proprietà rese disponibili che corrispondono con quelle richieste.
Questa tecnologia è ormai in uno stadio avanzato e sono molti gli speech engine che si trovano in rete, alcuni meno efficienti (voce robotica) ma gratuiti, altri molto simili ad un umano ma anche molto costosi.
Nel progetto viene sfruttata la parte front-end dello Speech Engine relativo alla sintesi vocale.
I sistemi attuali di sintesi vocale sono in grado di riprodurre con buona fedeltà la voce umana (anche se i suoni sembrano ancora metallici e la voce riprodotta non è ancora in grado di riprodurre correttamente le inflessioni e le emozioni umane) selezionando un particolare interlocutore (voce maschile, femminile, voce sexy,...).
Un esempio di questa tecnologia è presente nel sistema di sintesi vocale fornito da Microsoft nelle ultime versioni di Windows in lingua inglese (Microsoft Speech Engine). Basandosi su questo motore è stata costruita una serie di programmi in grado di parlare nelle varie lingue compreso l'italiano.

Vediamo in dettaglio lo schema di un sistema di sintesi vocale :

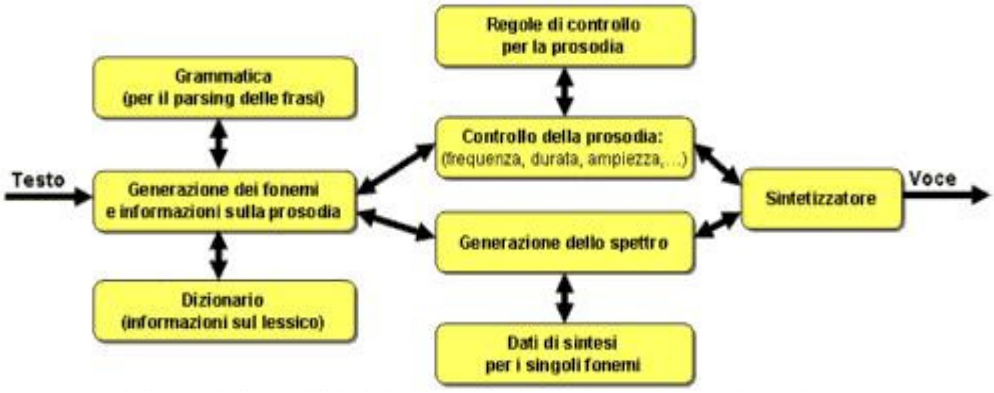

Figura 7 Schema di funzionamento di un sistema di sintesi vocale

Il testo da "leggere" viene fornito a un **generatore di fonemi** che basandosi su una **grammatica** e su un **dizionario** è in grado di scomporre le frasi in parole e le parole in singoli fonemi. Questa operazione, tutt'altro che semplice, deve tener conto di tutte le possibili eccezioni della lingua da riprodurre. Ad esempio i sistemi più vecchi pronunciavano la parola "aglio" senza pronunciare correttamente la "gl" e la parola "gnomo" senza pronunciare correttamente la "gn". Il parlato risultante aveva un forte accento americano. Grazie all'evoluzione delle tecnologie di analisi sintattica, grazie alla grammatica e al dizionario, i sistemi attuali non sono più affetti (tranne per qualche piccola eccezione) da questo spiacevole problema.

L'elenco dei fonemi viene trasferito a due moduli paralleli. Il primo è il **generatore dello spettro**. Il suo compito è quello di generare la forma d'onda risultante dall'unione di tutti i singoli fonemi basandosi su un archivio che contiene la rappresentazione sonora di tutti i fonemi (dati di sintesi). L'operazione non è semplice in quanto è necessario unire i vari suoni in modo da non introdurre pause indesiderate o suoni che una volta riprodotti sembrano a "scatti". L'altro modulo è il **controllo della prosodia**. Il suo compito è di inserire nella frase sintetizzata il "fattore emozione". Basandosi sulle **regole di controllo della prosodia**, il sistema è in grado di simulare le principali inflessioni della voce umana emulando anche alcune emozioni. Il risultato è una voce meno piatta e più "calda". Sembra quasi di sentire parlare un essere umano anche se con una qualità inferiore. Il risultato dell'elaborazione di questi due moduli viene passato ad un **sintetizzatore** che, utilizzando le normali schede audio presenti sui moderni PC emette la voce.

Dalla tabella che segue possiamo vedere le varie funzionalità di vari software di sintesi vocale (non tutti hanno un engine proprietario) e le caratteristiche che li differenziano l'uno dall'altro:

Software	Piattaforma compatibile	Conversione formati audio	Velocità della voce modificabile	Legge in ogni applicazione	API proprietario	Trial gratis
ClaroRead	Win	Mp3,Wav,Ogg		•		•
ReadPlease	Win	Mp3,Wav	•	•		•
NeoSpeech	Win	N/A		•	•	
TextAloud	Win	N/A	•	•	•	•
Loquendo	Win, Unix/Linux	Mp3,Wav	•	•	•	Demo web

Tabella 1. Tabella comparativa Software di sintesi vocale

Analizziamo in dettaglio le caratteristiche dei vari software (non tutti hanno delle API proprietarie per il text-to-speech per cui alcuni hanno solo l'interfaccia ma utilizzano le voci di altri software) identificando poi quello scelto per la realizzazione del progetto.

Claro Read

ClaroRead è un software estremamente efficace, una soluzione multisensoriale per tutti gli utenti che lottano con la lettura-scrittura a causa di dislessia o per altre situazioni. Progettato da esperti, è un software di qualità e di semplice utilizzo che permette ai suoi utenti la lettura di qualsiasi testo con una voce umana sia che si tratti di una pagina web, di un documento, di una e-mail o di qualsiasi altro testo si voglia. Esistono tre versioni: Standard, Plus e Premium e le lingue supportate sono inglese ,americano e per la versione Plus anche l'inglese indiano ma con un add-on si possono avere 17 lingue diverse.
Supporta i formati supportati sono Wma,Wav e Ogg.

La versione Standard di ClaroRead è integrata con Microsoft Word e Internet Explorer, rendendo così possibile la lettura di qualsiasi testo ad alta voce. Con ClaroRead Plus, si può ascoltare qualsiasi documento cartaceo dopo averlo opportunamente scannerizzato, convertire qualsiasi file PDF e riprodurlo con una pronuncia molto chiara.
Con ClaroRead Premium è possibile portare sempre con se il software di speech.Infatti è una versione appositamente progettata per le memorie USB portatili(i cosiddetti pennini USB).

ReadPlease

ReadPlease é un programma gratuito che permette al PC di leggere ad alta voce qualsiasi testo in Italiano e in varie lingue (Inglese, Tedesco, Francese, Spagnolo etc.), é anche disponibile una versione "Plus" a pagamento con maggiori funzioni. Permette inoltre la lettura di testi (modificandone anche la velocità) attraverso la clipboard di Windows (supporta copia/incolla), e di e-mail.E' possibile installare una ReadingBar per interagire direttamente col browser o con i documenti che si vogliono leggere ; utilizzando i controlli presenti su essa è possibile registrare i documenti in tempo reale e salvarli in formato Mp3 o Wav.
Per un uso professionale e per una migliore qualità delle voci é disponibile un set addizionale di 20 differenti "voci" della AT&T. Il programma puó essere anche molto utile per gli studenti o utenti disabili con dislessia, deficit visivo o ipovedenti.

NeoSpeech

La soluzione software NeoSpeech Voice Text è in grado di convertire il testo in suoni naturali ed è disponibile per vari device, desktop ed applicazioni network/server : è una delle migliori soluzioni presenti sul mercato. VoiceText supporta le lingue Inglese, Coreano, Giapponese e Cinese Mandarino e si avvale di una raccolta di undici voci native che sfruttano le lingue citate prima.

L'innovazione di questo specifico software è dovuta al fatto che è presente un dizionario base per ognuna delle lingue supportate, contenente centinaia di migliaia di pronunce ampliabile dagli sviluppatori, i quali possono inserirvi nuove pronunce di simboli, abbreviazioni e nuovi termini.
Inoltre è possibile adattare sia dinamicamente che per default i valori del pitch,della velocità e del volume del parlato. VoiceText tratta automaticamente input speciali come date, ore, abbreviazioni trovate in indirizzi e frasi con lingue miste.Supporta inoltre vari formati di output tra i quali l'audio 8kHz/16kHz sampling rates, linear 8-bit/16-bit PCM, 8-bit mu-law/a-law, ADPCM, .wav ed altri. Inoltre supporta SAPI5, C/C++, COM, e Java-based Application Programming Interfaces (APIs).

TextAloud

TextAloud è un sintetizzatore vocale che permette di trasformare il testo presente nell'interfaccia del programma in audio. Il software fa in modo che l'utente possa ascoltare il documento aperto invece di leggerlo. È possibile salvare il file vocale nei formati Mp3 e WMA per utilizzarlo su dispositivi portatili come iPod, Pocket PC e lettori Cd.
Tra le principali caratteristiche del programma: apertura di file Word, PDF e HTML, strumenti avanzati per la pronuncia , plug-in per inserire una toolbar in Internet Explorer, Mozilla Firefox e Outlook , pacchetti opzionali per gli accenti e le pronunce, filtri da applicare durante la sintetizzazione della voce(eco,riverbero,velocità, pitch etc.). E' compatibile con le voci Microsoft e da la possibilità di sfruttare le AT&T natural voices, le voci di NeoSpeech e quelle di Cepstral.

Loquendo

Un occhio di riguardo viene dato a quest'ultimo Speech Engine. Il motore di sintesi di Loquendo offre voci estremamente naturali, capaci di leggere qualunque tipo di dati e di messaggi nei servizi telefonici, nelle applicazioni multimediali, embedded e multimodali.
La metodologia di sintesi vocale è così affidabile ed efficiente da garantire a Loquendo una leadership di mercato in termini di qualità, portabilità, efficienza, naturalezza timbrica e intonativa e accuratezza di pronuncia.

Le voci Loquendo sono espressive, chiare, naturali e fluenti: sono state arricchite con un repertorio di termini ed eventi paralinguistici che permettono enunciati espressivi ed emozionali.

Loquendo TTS Director è un ambiente di sviluppo Java multi-piattaforma nato con l'obiettivo di agevolare gli utenti nella creazione di prompt per le loro applicazioni. I testi possono essere scritti direttamente all'interno della edit-box e immediatamente ascoltati, in modo da poter essere ulteriormente perfezionati fino al raggiungimento dell'effetto desiderato.

L'API Java di Loquendo TTSDirector permette di fissare i parametri acustici e prosodici (ad esempio la frequenza di campionatura, la tonalità, la velocità di lettura e il volume) e salvare il prompt creato sia in formato testo, che in formato audio ma è possibile interfacciarsi al tts utilizzando C Loquendo API (Win32, Linux32), C++ Loquendo API (Win32),SAPI4e5(Win32),W3C SSML1.0,Loquendo TTS ActiveX (Win 32).

Loquendo TTS offre i più alti livelli di flessibilità, scalabilità, performance e robustezza : la propria configurazione *multi-thread* e *multi-processo* permette lo sviluppo di applicazioni in qualunque architettura software e di soddisfare ogni requisito tecnico e commerciale.

Loquendo TTS implementa un algoritmo molto accurato ed efficiente che garantisce una risposta estremamente rapida; il software vocale può sintetizzare differenti lingue e voci simultaneamente, passando da una all'altra in qualunque momento su qualunque canale.

Il lessico di pronuncia assicura che i vocabolari specialistici, le abbreviazioni, gli acronimi e le flessioni regionali siano realizzate secondo l'intenzione dello sviluppatore e vengono pronunciati correttamente anche i formati speciali, quali numeri telefonici, valute e indirizzi e-mail oltre al fatto che le caratteristiche di ciascuna voce (ad esempio il tono, la velocità di eloquio e il volume) possono essere ottimizzate e controllate in ogni aspetto.

Le tecnologie Loquendo vengono proposte in 20 lingue, tra le quali: Inglese e Americano, Francese per il Canada, Castigliano, Catalano, Valenziano, Messicano, Cileno e Argentino, Latino Americano, Italiano, Tedesco, Francese, Greco, Cinese Mandarino, Olandese, Brasiliano, Portoghese e Svedese ,sia nella voce femminile che in quella maschile.

Loquendo TTS è disponibile nelle versioni Telefonica, Multimedia e Embedded, garantendo lo stesso ampio spettro di voci e lingue e lo stesso algoritmo nei vari ambienti.

In conclusione dotando il sistema di una implementazione dello speech engine è possibile sfruttare le sue funzioni di sintesi e riproduzione vocale. In questo modo, come è stato mostrato, è semplice inserire le potenzialità che questa tecnologia offre nelle proprie applicazioni, per rendere più user-friendly le interfacce utenti.

2.2 Avatar animati

Un altro campo, che si è sviluppato molto negli ultimi anni, è quello degli avatar animati.
Navigando ci si rende conto che si è ormai sommersi da innumerevoli immagini, icone e soprattutto avatar atti a rappresentare l'utente che ne fa uso. Il termine avatar ha origini lontanissime (deriva da avatara) che designa le diverse incarnazioni degli dei in India. Per intenderci, Buddha è un avatara. Esiste, dunque, un curioso legame tra una delle religioni più antiche del mondo e una delle più avanzate tecnologie applicate all'informatica e alla telematica.
Gli avatar sono delle "incarnazioni" virtuali dei loro creatori. Una sorta di trasposizione in qualcosa di ultra terreno, visibile su spazi tridimensionali all'interno dei quali agiscono. Tutto nasce dall'abilità del lavoro di un buon programmatore che decida di "incarnarsi" in un essere virtuale. Dal punto di vista grafico, un avatar è rappresentato da un mezzo busto all'interno di un rettangolo che ne definisce il confine visivo. Con opportuni comandi è possibile far muovere il collo, le spalle, gli occhi, il mento, la bocca, il naso e le sopracciglia.
Oggigiorno è possibile quindi costruire il proprio avatar fino a raggiungere livelli di perfezione strabilianti ; questi ormai ci rappresentano alla perfezione,sono animati, ci guardano, generano emozioni:ci somigliano!
Nella realizzazione dei software capaci di emulare una faccia animata devono essere implementati moltissimi aspetti come lo specifico movimento delle labbra in riferimento ai vari fonemi pronunciati , la sincronizzazione del parlato e del movimento del viso , il modo in cui viene effettuato il TTS ecc..
Analizziamo quindi i migliori software che sono attualmente disponibili in commercio e che permettono di creare avatar animati :

Software	Photo Import	Morphing	Controllo con Javascript
Oddcast	No	No	Yes
Kallideas	No	No	No
H-Care	No	No	No
Microsoft Agent	No	No	Yes
CrazyTalk	Yes	Yes	Yes

Tabella2. Tabella Comparativa software per Web Avatar

Oddcast

Oddcast è un ottimo sintetizzatore vocale molto semplice da utilizzare. Accedendo alla home page si fa la conoscenza di una simpatica signora che passandole il mouse sul voto inizierà a muovere gli occhi verso il puntatore con fare bizzarro. In basso, invece, si troverà la casella dove inserire il testo da farle pronunciare.

Tra le lingue supportate da Oddcast ,oltre all'italiano, è presente l'inglese, il francese, il tedesco, l'americano, lo spagnolo, il cinese e tantw altre. Con riferimento alla lingua italiana,pronunciata piuttosto bene, è possibile scegliere tra ben 9 personaggi, ognuno con cadenza e tonalità diversa.

Tra i prodotti Oddcast, quello più potente è VHost Studio che comprende un software di TTS per dare la voce agli avatar creati ; è uno strumento di authoring facile da usare che permette agli utenti anche non esperti in fatto di programmazione web, di creare e inserire avatar animati adattati all'interno di pagine HTML, banners o filmati FLASH. Tutto questo infatti è realizzato tramite un semplice tool che guida l'utente alla creazione del web avatar con pochi semplici step.

I programmatori HTML e JavaScript possono utilizzare le API di VHost Studio per creare interazioni avanzate con gli utenti basandosi sui loro rollovers, sui loro clicks e sui cookies del browser : le funzionalità TTS permettono al VHost di pronunciare qualsiasi testo in maniera dinamica, in tempo reale, con sincronizzazione labiale precisa.

Queste funzionalità rappresentano un'alternativa, o meglio un completamento delle funzionalità esistenti di VHost , che permettono agli utenti di caricare audio pre-recorded (formato Wav o Mp3) affinché l'avatar realizzato lo riproduca.

Il VHost TTS ha un insieme di API, che permettono di collegarlo a qualsiasi database, per dare esperienza di carattere conversazionale effettivamente dinamica. Unici nei la mancanza della funzionalità di import di immagini personali e di morphing delle stesse.

Kallideas

Kallideas si propone nel mondo degli assistenti virtuali con un progetto chiamato K-Humans.
Il loro scopo è rendere più amichevole ed efficace l'interazione fra utenti e sistemi automatici, per permettere alle aziende che li adottano di dialogare in modo efficiente e innovativo con clienti, prospect, dipendenti ecc.. Gli Assistenti Virtuali sono un sogno tecnologico che esiste da almeno venti anni e che è stato possibile realizzare soltanto di recente ; Questi VA (virtual assistants) sono il risultato dell'applicazione dei Sistemi Decisionali Esperti (Intelligenza Artificiale) alle estese Knowledge Base dei contact center aziendali e progressivamente a quelle dei contenuti leisure ed entertainment, il tutto fuso poi con un Engine Grafico di nuova concezione che genera e muove una persona virtuale consentendole di parlare e soprattutto di comprendere e interpretare le domande poste, parlando o scrivendo, dal suo interlocutore umano con una modalità definita HumanLike interaction.

Oltre a queste caratteristiche, i K-humans realizzati da Kallideas sono dotati di un particolare software (BRAIN) che dona ad essi una capacità specifica di relazionarsi in modo emotivamente connotato : i sistemi sono capaci di intuire se l'interlocutore umano è in difficoltà nella comprensione dei processi che sta chiedendo di gestire ed adottare un approccio più didattico.
I K-humans possono assumere qualsiasi "ruolo" aziendale: sportellista o consulente in una banca, presso punti informativi di ogni genere, nella didattica a distanza, al centro prenotazioni di un ospedale, al check-in di un aeroporto o di un call-center.

In definitiva un K-humans assiste i clienti nel:

- Ricercare informazioni
- Esprimere i propri bisogni
- Interagire con altri sistemi automatici (ad es. prenotazioni, portali web, etc…)

e interagisce con loro utilizzando: Voce (telefono, web,), messaggi di testo, gestualità, video e immagini.

H-Care

H-care sviluppa soluzioni tecnologiche innovative per i servizi di customer care in modalità self-service e multimodale : uno dei vari prodotti sviluppati è la piattaforma software Human Digital Assistant, che consente di realizzare un operatore digitale di customer care disponibile tramite diversi canali di interazione con la clientela combinando all'alta qualità dell'animazione 3D in tempo reale , una sintesi vocale all'avanguardia integrandosi liberamente con l' infrastruttura esistente per ottenere un assistente dinamico human-like.

La piattaforma HDA è composta da una serie di componenti e moduli che danno vita al Digital Assistant garantendo una efficiente comunicazione tra utenti e servizi:

- **Brain Server** : Brain è il motore che gestisce le interazioni e la logica dietro la faccia dell'assistente. Brain è un applicazione java server-side.
- **Environment** : L'architettura del plugin Brain permette più collegamenti tra l'assistente e sistemi esterni come database relazionali, CRM, portali ed altri sistemi di impresa. Un software di sviluppo SDK (java-based) permette di costruire velocemente e personalizzare i dati esistenti.
- **Face Module (Server-side)** : comprende animazioni e rendering 3D di alta qualità,un text-to-speech per la generazione della voce dell'assistente e la codifica audio/video in differenti formati. Il face engine è il nucleo nonché il componente più innovativo della piattaforma HDA capace di dare il più avanzato carattere human-like a qualsiasi dispositivo in rete e il tutto in real-time.
- **HSC (Client-side)** : L'HDA smart client gestisce tutti gli aspetti relativi alle interazioni con l'utente , è un modulo lato client che deve essere integrato nella web application di destinazione e la personalizza utilizzando un SDK.

Microsoft Agent

L'agente Microsoft è stato il primo avatar parlante utilizzato nei software di casa Microsoft come help per gli utenti ; abbinandolo ad un Text-To-Speech o assegnandoli un audio pre-recorded, esso può essere utilizzato anche in applicazioni o su pagine web.

L'agente Microsoft è non solo versatile dal punto di vista dell'utilizzo ma anche dal punto di vista dello sviluppo dato che si possono scegliere dei tools di vari linguaggi che supportano le tecnologie ActiveX come il Microsoft Visual Basic , i sistemi di sviluppo in Visual C++ e molti altri.

Gli autori Web possono incorporare velocemente Microsoft Agent nelle loro pagine HTML utilizzando Visual Basic Scripting Edition (VBScript) e software di sviluppo JScript e in tutte le altre applicazioni supportano Visual Basic Application (VBA) come Microsoft Office.

CrazyTalk

Confrontandolo con gli altri software presenti in commercio , CrazyTalk della Reallusion , che illustreremo nei paragrafi successivi, si è rivelato il più efficiente e completo prodotto per la creazione e la configurazione di web avatar.

2.2.1 Il plug-in Crazy Talk

Le interfacce multimodali sono inserite all'interno delle pagina web come plug-in.
Il plugin (o plug-in) è un programma non autonomo che interagisce con un altro programma per ampliarne le funzioni. Il tipico esempio è un plugin per un software di grafica che permette l'utilizzo di un formato grafico non supportato in maniera nativa dal software principale. La capacità di un software di supportare i plugin è generalmente un'ottima caratteristica rendendo così possibile l'ampliamento delle sue funzioni in maniera semplice e veloce.

Il plugin utilizzato in questo contesto per realizzare l'interfaccia multimodale all'interno di una pagina web, è stato sviluppato dalla società Reallusion e prende il nome di Crazy Talk (CT) ; può essere scaricato gratuitamente presso il sito web del produttore e ha dimensioni molto ridotte (circa 1Mb). La potenza di questo prodotto risiede soprattutto nella semplicità di creazione e di configurazione di tali web avatar all'interno di una qualsiasi pagina web : una volta fornita la foto del personaggio da animare e dopo aver selezionato correttamente i contorni degli elementi principali costituenti il volto, Crazy Talk è in grado di animare in maniera del tutto naturale la faccia, imprimendo così espressioni facciali sempre diverse.
Il plug-in si compone di due parti: il crazytalk web player e il crazytalk TTS. Il primo consente di effettuare l'animazione del personaggio scelto, utilizzando opportuni file creati preventivamente con il software CT in dotazione, il secondo di produrre l'audio fruttando il suo speech-engine.

Crazy Talk Web Player

Il CrazyTalk Web Player può essere inserito in una pagina web utilizzando il codice specificato sotto.

```
<OBJECT ID="CrazyTalk" classid="CLSID: 13149882-F480-4F6B-8C6A-0764F75B99ED"
codebase="http://plugin.reallusion.com/CrazyTalk4.cab#version=1,0,0,0" width="150"
height="200">
<PARAM Name="ModelName" Value="crazytalk.ctm">
<PARAM Name="ScriptName" Value="crazytalk.cts">
<PARAM Name="ControlStyle" Value="1">
<PARAM Name="BorderStyle" Value="1">
<PARAM Name="LifeMode" Value="1">
<PARAM Name="AutoPlay" Value="1">
</OBJECT>
```

Il tag OBJECT consente di inserire all'interno delle pagine web file multimediali (audio e video), oppure effetti grafici particolari. Gli attributi principali di questo tag sono:

nome attributi	descrizione
id	il nome del file multimediale inserito
classid	dà indicazioni sul percorso dell'oggetto, ed è utile per identificare il tipo di plugin con cui eseguire l'oggetto
codebase	indica l'URL e la versione del player
width	indica la larghezza della finestra contenente il player
height	indica l'altezza delle finestra contenente il player

Tabella 3 - Attributi del tag object -

Il tag PARAM specifica le proprietà del player inserito all'interno della pagina web.
Sono di notevole importanza i parametri "ModelName" e "ScriptName" che indicano i file che devono essere presi in considerazione per l'animazione del viso:
- .CTM è il modello della faccia creato con il software CrazyTalk 4.5 che permette di settare tutti i parametri del viso come colore degli occhi, la forma delle labbra e in generale tutti i punti per la modellazione della faccia.
- .CTS descrive il file audio da riprodurre e contiene tutte le informazioni sulla sincronizzazione parlato/movimenti del viso.

Il plugin crazytalk, inserito all'interno della pagina, è caratterizzato da:

- proprietà;
- metodi;

Le prime consentono di modificare le caratteristiche visive e sonore della faccia animata e sono specificate nei campi del tag param; i secondi vengono utilizzati per modificare le proprietà.
Inoltre, esistono dei particolari messaggi che vengono inviati dal crazytalk web player al browser in presenza di alcune particolari azioni. Questi messaggi vengono chiamati eventi. È possibile catturarli e manipolarli a piacimento. Per una trattazione dettagliata rimandiamo all'appendice B.

Crazy Talk TTS

Il crazytalk TTS è un oggetto che può essere inserito all'interno delle pagine web esattamente come nel caso del crazytalk Web Player. L'utilizzo di questo plugin consente di:

1. visualizzare il testo che il web player sta producendo come audio (quindi come dei sottotitoli);
2. comandare il web player direttamente sul client;

La visualizzazione del testo consente di migliorare le prestazioni del crazytalk, in quanto permette all'utente di seguire lo script audio all'interno della casella testuale che realizza la grafica del plugin. I vantaggi di questa soluzione sono molti: ad esempio in presenza di un termine audio poco chiaro è possibile utilizzare la versione testuale per decifrarlo, oppure se il terminale utente è sprovvisto di una scheda audio si può pensare di utilizzare il testo come soluzione temporanea.

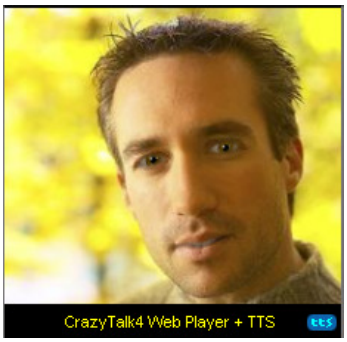

Figura 8 CrazyTalk TTS (Text-To-Speech)

L'utilizzo principale che si può fare, però, del TTS è quello di comandare dal browser, tramite funzioni javascript, il comportamento del player che è ciò che ci interessa maggiormente per i nostri scopi.

Gli script audio (e le relative espressioni facciali) possono essere, infatti, generati in due modi:

- staticamente;
- dinamicamente;

Nella prima soluzione si utilizza il software Crazy Talk 4.5 : definito il testo da riprodurre, l'applicazione genera il file con estensione .cts che deve essere inserito all'interno della proprietà ScriptName.

La seconda soluzione, invece, consente di generare dinamicamente il testo in funzione delle operazione svolte dall'utente. Non serve pertanto creare un file .cts tramite il software dato che il file audio verrà generato automaticamente tramite lo speech engine indicato (insieme ad altri parametri) nel PARAM "TTSEngine" dell'oggetto OBJECT relativo al TTS PLayer.
Ad esempio si può utilizzare per far ripetere all'interfaccia i valori inseriti in fase di registrazione.
In sostanza la prima soluzione crea degli script fissi, che non variano da utente a utente. La seconda soluzione può essere utilizzata per far ripetere all'interfaccia termini inseriti dall'utente in occasione di specifici eventi generati.
Ovviamente in entrambe le situazioni ci sono svantaggi e vantaggi:

soluzione statica	soluzione dinamica
lo script è prodotto staticamente sul server e non può variare	lo script varia dinamicamente in funzione delle operazioni svolte dall'utente
lo script è costituito da un file con estensione .cts e prima di poterlo utilizzare deve essere scaricato dal server	lo script viene generato sul client (anche in stato offline) e non occupa, quindi, in alcun modo la connessione.
per poter limitare i tempi di risposta dell'interfaccia le dimensioni dello script non possono essere elevate	lo script può essere di qualsiasi dimensione
lo script è indipendente dalla configurazione del browser utilizzato dal client	per funzionare correttamente il browser utente deve avere abilitato l'uso del javascript
per generare il file .cts il software Crazy Talk ha bisogno di un TTS engine installato solo sulla macchina su cui si effettua lo sviluppo del sito web	il TTS engine deve essere installato su tutti browser

Tabella 4 – Vantaggi e svantaggi del crazytalk TTS -

Come si può notare anche dalla tabella 2, la principale limitazione del crazytalk TTS è la necessità di avere un TTS engine installato sulla macchina dove vengono prodotti gli script, indipendentemente dalla modalità di creazione (dinamica o statica). In commercio sono molti i TTS engine disponibili, alcuni gratuiti, altri a pagamento. I più comuni sono: microsoft SAPI 5.0 (gratuito) e Loquendo TTS (a pagamento).

Il codice necessario ad inserire il crazytalk TTS all'interno di una pagina web è simile a quello già analizzato per il crazytalk web player.

```
<OBJECT ID="RLTTSPlayer" classid="CLSID: B7A59580-B39D-4BF9-B968-
1BFA25156691"codebase="http://plugin.reallusion.com/CrazyTalk4.cab#version=1,0,0,0"
width="256" height="24">
        <param name="BackColor" value="16711680">
        <param name="ForeColor" value="65535">
        <param name="FontName" value="Arial">
        <param name="FontSize" value="100">
        <param name="TTSEngine" value="Microsoft SAPI 4.0">
        <param name="Speed" value="50">
        <param name="Pitch" value="50">
        <param name="Volume" value="80">
        <param name="AutoLoop" value="0">
        <param name="TextContent" value=" ……">
TTS Engine<br>
Please install TTS Web Player to see the Talking Head.<br>
<a href="http://www.reallusion.com/plug-in/installct.asp">
http://www.reallusion.com/plug-in/installct.asp</a>
</OBJECT>
```

Come nel caso del crazytalk web player anche per il TTS vengono definite proprietà, metodi ed eventi. Per una descizione dettagliata vedere APPENDICE B.

Tra le proprietà fondamentali possiamo menzionare:

- TextContent nel quale viene specificato la stringa di testo da convertire tramite il TTS.
- TTSEngine specifica il TTS da utilizzare per effettuare la conversione.

Per poter utilizzare la funzionalità di creazione dinamica degli script occorre collegare il TTS con il web player. Per questo è necessario inserire all'interno delle pagine il seguente codice:

```
<SCRIPT language="javascript">
"id_TTS".AttachCtrl("id_player");
</SCRIPT>
```

Dove i termini id_TTS e id_player sono i valori inseriti nel campo id del tag object.

Un esempio completo quindi puo' essere:

```
//inserimento oggetto crazytalk player
<OBJECT CLASSID="CLSID: 13149882-F480-4F6B-8C6A-0764F75B99ED"
id="CrazyTalkPlayer" width="256" height="256">
  <param name="ModelName" value="Luka.ctm">
  <param name="ScriptName" value="Luka.cts">
</OBJECT>

//inserimento oggetto TTS
<OBJECT CLASSID ="CLSID: B7A59580-B39D-4BF9-B968-1BFA25156691"
id="RLTTSComponent" width="256" height="24">
  <param name="TTSEngine" value="Microsoft SAPI 5.0">
</OBJECT>
// link fra TTS e player
<SCRIPT language="javascript">
RLTTSComponent.AttachCtrl(CrazyTalkPlayer);
</SCRIPT>
```

2.3 Il linguaggio lato client : Javascript

Un ultimo, ma non in termini di importanza, strumento da tenere in considerazione per la realizzazione di una IM è un certo linguaggio che ci permetta di far svolgere a quest'ultima, specifiche azioni in relazione a eventi generati dal suo utilizzatore.

Quando visualizziamo le nostre pagine web da casa ci sono due computer che si parlano: il server ed il client. Alcuni linguaggi di scripting (asp,php) vengono eseguiti dal web server (si chiamano appunto linguaggi server side o lato server) mentre altri vengono eseguiti sul nostro computer di casa dal browser (i cosidetti linguaggi lato client) ; il piu' potente tra questi ultimi è **Javascript**.

E' un linguaggio apparentemente semplice da imparare perchè è molto simile come struttura e sintassi ad altri linguaggi di programmazione come il C++ e il Java (dispone infatti di funzionalità orientate agli oggetti).

JavaScript permette l'inserimento di contenuti eseguibili all'interno di pagine web, permette così la creazione di pagine HTML dinamiche interagendo con l'utente, controllando il browser e creando nuovi contenuti HTML.

Con Javascript sarà possibile effettuare chiamate ai metodi dell'interfaccia di web avatar utilizzata (nel nostro caso quella fornita da CrazyTalk) in relazione ad eventi generati dall'utente ; naturalmente il web avatar scelto deve mettere a disposizione al programmatore una specifica API Javascript utile a tal fine.

In questo modo sarà possibile ad esempio far pronunciare una specifica frase alla nostra IM, o muovere gli occhi, le spalle e così via, in conseguenza ad un evento onclick (ad esempio click su un campo di inserimento testo) oppure onmouseover (passaggio con il mouse su una immagine i su un link) , o mouseout (pressione di un tasto del mouse) , onload (caricamento della pagina web) e così via.